ASCO 2023 — Poster 249, Abstract 3051, June 3, 2023

Detection of early-stage cancers using circulating orphan non-coding RNAs in blood

Mehran Karimzadeh[1], Jeffrey Wang[1], Taylor B. Cavazos[1], Lee S. Schwartzberg[2], Michael Multhaup[1], Jeremy Ku[1], Xuan Zhao[1], Jieyang Wang[1], Kathleen Wang[1], Rose Hanna[1], Patrick Arensdorf[1], Kimberly H Chau[1], Helen Li[1], Hani Goodarzi[3], Lisa Fish[1], Fereydoun Hormozdiari[1], Babak Alipanahi[1]

[1]Exai Bio Inc., Palo Alto, CA; [2]Renown Health-Pennington Cancer Institute, Reno, NV; [3]University of California San Francisco, San Francisco CA

Background

- Orphan non-coding RNAs (oncRNAs) are a novel category of small RNAs (smRNAs) that are frequently detected in cancer and largely absent in non-cancerous tissues.
- First identified in breast cancer samples from The Cancer Genome Atlas (TCGA)[1], novel oncRNAs have, since then, been discovered in additional cancer tissues from TCGA and validated in an independent cohort of tumor and adjacent normal tissues[2].
- We recently assessed the oncRNA content of serum and demonstrated their potential to detect colorectal[3], breast[4], and lung[5] cancers in a liquid biopsy strategy. These investigations, however, have been limited to single cancer cohorts, and the broader applicability of oncRNAs as biomarkers in a single multi-cancer blood test has yet to be determined.
- In this study, we investigate the utility of oncRNAs as serum biomarkers for early cancer detection across eight cancer types.

Goals

- Develop and validate an artificial-intelligence (AI)-driven blood test for cancer detection in multiple cancers across a range of cancer stages.
- Build and evaluate a cancer tissue-of-origin classification model using each patient's serum-oncRNA profile.

Methodology

- We collected 3,317 serum samples from individuals with known cancers of the bladder (*n* = 164), breast (*n* = 220), colon and rectum (*n* =143), kidney (*n* = 293), lung (*n* = 295), prostate (*n* = 96), pancreas (*n* = 346), and stomach (*n* = 286), as well as donors with no history of cancer at time of collection (*n* = 1,474).
- Patients had provided informed consent and contributing centers had obtained IRB approval.
- We used 0.5 mL serum aliquots to generate and sequence smRNA libraries at an average depth of 20 million 50-bp single-end reads. Individuals were split into age-, sex-, and smoking status-matched training (1,377 cancer; 1112 control) and test (466 cancer; 362 control) sets. We then profiled all the serum samples with our catalog of TCGA-derived oncRNAs[2].
- We trained generative AI models with batch effect removal, library-size estimation, and expression normalization modules to predict cancer presence and tissue-of-origin (TOO) through five-fold cross-validation within the training set. For individuals with cancer and high-confidence AI prediction, we also reported TOO.
- We evaluated the generalizability of our model by predicting cancer status and TOO in the held-out test set. Predictions were averaged across the five models optimized on the training set folds.

Overview of oncRNA Profiling and AI-Driven Model for Cancer Prediction

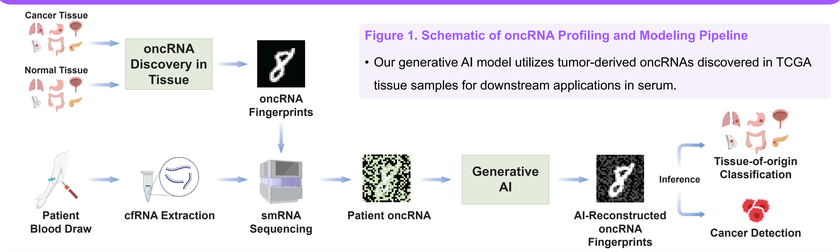

Figure 1. Schematic of oncRNA Profiling and Modeling Pipeline

- Our generative AI model utilizes tumor-derived oncRNAs discovered in TCGA tissue samples for downstream applications in serum.

Study Demographics

Characteristics		Training Set		Test Set	
		Cancer	Control	Cancer	Control
Total (*n*)		1,377	1,112	466	362
Age (mean, SD)		62.3 (12.2)	58.6 (12.8)	62.2 (11.6)	59.3 (11.5)
Sex (*n*, %)	Female	652 (47.4%)	539 (48.5%)	209 (44.8%)	190 (52.5%)
	Male	725 (52.7%)	573 (51.5%)	257 (55.2%)	172 (47.5%)
Cancer Stage (*n*, %)	I	507 (36.8%)	–	165 (35.4%)	–
	II	437 (31.8%)	–	152 (32.6%)	–
	III	254 (18.5%)	–	96 (20.6%)	–
	IV	179 (13.0%)	–	53 (11.4%)	–
Cancer Diagnosis (*n*, %)	Lung	221 (16.0%)	–	74 (15.9%)	–
	Gastric	215 (15.6%)	–	71 (15.2%)	–
	Pancreas	255 (18.5%)	–	91 (19.5%)	–
	Kidney	220 (16.0%)	–	73 (15.7%)	–
	Colorectal	106 (7.7%)	–	37 (7.9%)	–
	Breast	165 (12.0%)	–	55 (11.8%)	–
	Prostate	72 (5.2%)	–	24 (5.2%)	–
	Urothelial	123 (8.9%)	–	41 (8.8%)	–

Result 1: Ability of oncRNA-Based Model for Prediction of Overall Cancer Status in Serum

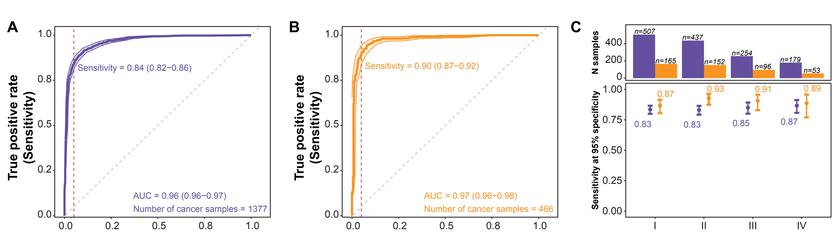

Figure 2. Overall Model Performance by ROC and Sensitivity at 95% Specificity

- **(A)** The ROC curve demonstrated an AUC of 0.96 (95% CI: 0.96-0.97) in the training set, and **(B)** an AUC of 0.97 (95% CI: 0.96-0.98) in the test set. AUC 95% confidence intervals were calculated by bootstrapping.
- **(C)** Sensitivities at 95% specificity for the training set (purple) and test set (orange) stratified by tumor stage. Confidence intervals report the 95% confidence intervals calculated using the Clopper-Pearson method within each group. Bar plots in the top panel show the number of samples corresponding to each tumor stage and training/test set.

Result 2: Discrimination Between Cancer Patients and Cancer-Free Controls in Serum Across Cancer Diagnoses

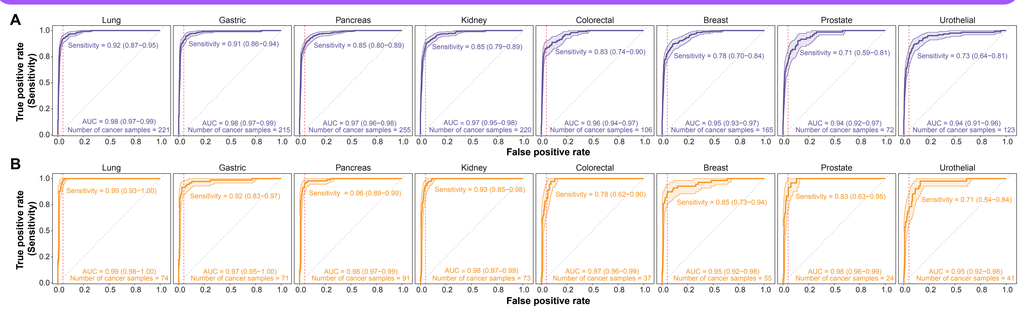

Figure 3. Model Performance By Cancer

- Our model had high accuracy (AUC ≥ 0.94), demonstrating robust prediction across eight cancer types, within both the training **(A)** and held-out test **(B)** sets.
- AUCs ranged from 0.95 (95% CI: 0.92–0.98) in urothelial cancer to 0.99 (95% CI: 0.98–1.00) in lung cancer within the test set **(B)**.
- Sensitivities at 95% specificity were lowest for urothelial cancer (0.71, 95% CI: 0.54–0.84) and highest for lung cancer (0.99, 95% CI: 0.93–1.00) in the test set **(B)**.

Result 3: Model Accurately Predicts Tumor Tissue-of-Origin for Cancer Patients within Serum

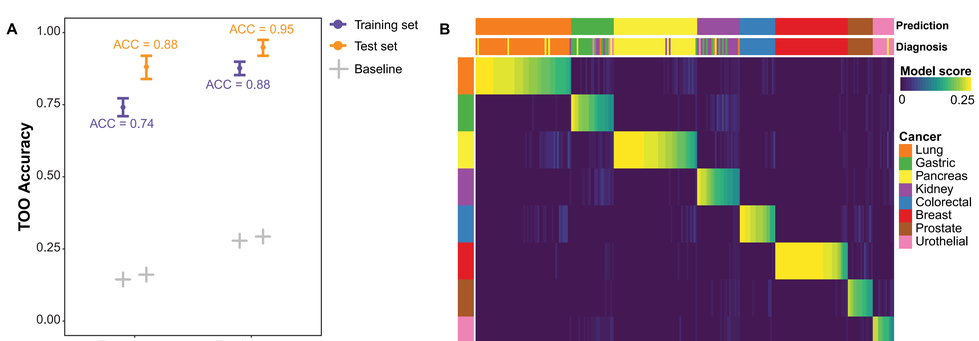

Figure 4. Performance and Score Distribution of the Tissue-of-Origin Model

- **(A)** For samples with cancer tissue-of-origin prediction, our held-out testing cohort had an accuracy of 0.88 (95% CI: 0.84-0.92) using the top predicted cancer and 0.95 (95% CI: 0.92-0.97) for the top two predictions. 95% CIs were computed through bootstrapping. The baselines shown in gray represents the expected performance of a random guess given cancer prevalence within our study. **(B)** The heatmap shows the predicted scores of the model for each sample and cancer type.

Conclusions

- Our results show that circulating serum oncRNAs captured through a liquid biopsy assay can be used to accurately detect a shared cancer signal in a single, multi-cancer early detection test.
- We also demonstrate that a multi-modal generative AI model trained on oncRNA profiles can robustly and accurately predict cancer tissue-of-origin within serum for samples with detectable cancer signals.
- Given the limitations of retrospective studies and the use of frozen, archival samples, we plan to further validate our results in the future through a larger, prospective study.

Disclosures: MK, JW , TC, MM, JK, XZ, JW, KW, RH, KC, HL, LF, and FH are full-time employees of Exai Bio. BA and PA are co-founders, stockholders, and full-time employees of Exai Bio. HG is a co-founder, stockholder, and advisor of Exai Bio. LS is an unpaid advisor of Exai Bio.

References:

1. Fish L, et al. *Nature Med.* 2018;24:1743-51.
2. Wang J, et al. AACR. 2022; 3353.
3. Wang J, et al. ESMO. 2022; 4635.
4. Cavazos T, et al. SABCS. 2022; P1-05-18.
5. Karimzadeh M, et al. AACR 2023; 5711.